View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Caltech Authors

# Supplementary Methods

## Stimulation

The following reports in detail the stimuli used in the various experiments. We define the stimulus variable $s(t)$ as the intensity $I(t)$ normalized for mean and contrast:

(5) $$s(t) = \left(I(t) - M\right)/C\,.$$

Also, we define the random variable $n_{15}(t)$ as a staircase waveform updated every 15 ms by independent draws from a normal distribution; similarly $n_{30}(t)$ is updated every 30 ms.

**Spatial correlation (Fig 1a):** The field was divided into two sets of alternating square tiles like a checkerboard (Fig S1a). One set was modulated with stimulus variable $x(t)$, the other with $y(t)$. Environment A (positive correlation): $y(t) = x(t) = n_{30}(t)$. Environment B (negative correlation): $y(t) = -x(t) = n_{30}(t)$. P (probe): $y(t)$ and $x(t)$ are independent versions of $n_{30}(t)$. The tile size was chosen as 400 $\mu$m or 200 $\mu$m, similar to the diameter of a typical receptive field center for salamander ganglion cells.

**Spatial orientation (Figs 2a):** The field was divided into square tiles belonging to 4 symmetrical sets (Fig S1b) modulated with stimulus variables $x(t), y(t), u(t), v(t)$. A (horizontal bars): $x(t) = y(t) = -u(t) = -v(t) = n_{30}(t)$. B (vertical bars): $x(t) = -y(t) = u(t) = -v(t) = n_{30}(t)$. P (probe): $x(t), y(t), u(t), v(t)$ all modulated independently as $n_{30}(t)$. The tile size was chosen as 200 $\mu$m. For the "shifting border"

condition (Fig 2e), the tiling was shifted randomly on a fine square grid (40 or 67 $\mu$m) at every stimulus update (30 ms).

**Temporal correlation (Fig 3a):** A uniform field was modulated with stimulus variable $x(t)$. A (positive correlations across 60 ms):

$x(t) = 0.97 \cdot x(t - 60\text{ ms}) + \sqrt{1 - 0.97^2} \cdot n_{15}(t)$. B (negative correlations across 60 ms):

$x(t) = -0.97 \cdot x(t - 60\text{ ms}) + \sqrt{1 - 0.97^2} \cdot n_{15}(t)$. P (uncorrelated probe): $x(t) = n_{15}(t)$.

**Space-time delay (Fig 3e):** The field was again divided like a checkerboard with 200 $\mu$m tiles (Fig S1a), and the two sets of tiles modulated with stimulus variables $x(t)$ and $y(t)$. A (X advanced): $y(t) = x(t - 60\text{ ms}) = n_{30}(t - 60\text{ ms})$. B (Y advanced): $x(t) = y(t - 60\text{ ms}) = n_{30}(t - 60\text{ ms})$. P (probe): $x(t), y(t)$ drawn independently as $n_{30}(t)$.

## Analysis

To characterize each neuron's light response, the brief probe segments P were used to derive the best-fit LN model [37] for the firing rate. For example (Figs 1, S2), if the stimulus contains two spatial regions modulated with $x(t)$ and $y(t)$, then the LN fit to the firing rate is

$$r(t) = N(g(t)) = N\left(\int x(t')L_X(t - t')dt' + \int y(t')L_Y(t - t')dt'\right). \quad (1)$$

The linear filter was obtained from the first-order kernel of the spike train with respect to the stimulus variables.

$$L_X(\tau) = \frac{1}{T}\int_0^T x(t - \tau)r(t)dt$$
$$L_Y(\tau) = \frac{1}{T}\int_0^T y(t - \tau)r(t)dt \quad (6)$$

where $T$ is the duration of the spike train used for analysis. Then the nonlinearity $N(g)$ was found by computing

$$g(t) = \int x(t')L_X(t-t')dt' + \int y(t')L_Y(t-t')dt' \tag{7}$$

and plotting the measured firing rate $r(t)$ against $g(t)$.

In computing the filter, only the first 0.8 s of each P segment were used, to limit the degree of adaptation to the probe stimulus itself. Moreover, the very beginning of each P segment, amounting to the duration of the filter (typically 0.18 s), was ignored to avoid contamination of the response from the preceding adapting stimulus. Note that this also precluded the detection of any fast changes in the neuron's sensitivity that may occur instantaneously on switching from environments A or B to P. Such very rapid changes are observed, for example, when the switch involves a simple change in stimulus contrast [13].

Two models were computed for the adapting conditions A and B. For any given cell, the shape of the nonlinearity was found to be essentially the same under both A and B, and thus we used the same function $N(g)$ in fitting the model to both conditions. The resulting LN model produced a good fit to the recorded spike trains, with an RMS deviation between model and neuron [37] of typically 0.15 ± 0.01 spikes per 15 ms time-bin (mean ± SEM, 15 cells, experiment of Fig 1).

To assess the degree of predictive coding, we evaluated how sensitive the neuron is to stimuli drawn from environments A and B. Specifically, we measured the root-mean-square amplitude of the output $g(t)$ from the linear filter (Fig S2), if it were stimulated with ensemble A or B. For example, consider the spatial correlation experiment (Fig 1, Eqn (1)). In environment A, $y(t) = x(t) = n_{30}(t)$, and therefore

(8)
$$\begin{aligned} g(t) &= \int x(t')L_X(t-t')dt' + \int y(t')L_Y(t-t')dt' \\ &= \int n_{30}(t')\left[L_X(t-t') + L_Y(t-t')\right]dt' \end{aligned}$$

Because the values of $n_{30}(t')$ are drawn independently from a normal distribution, the variance of $g(t)$, averaged over all stimuli in A, is to within some constant factor

(9)
$$\langle g^2 \rangle_A \propto \int \left[L_X(t') + L_Y(t')\right]^2 dt' .$$

In environment B, $y(t) = -x(t) = n_{30}(t)$, and therefore

(10)
$$g(t) = \int n_{30}(t')\left[L_X(t-t') - L_Y(t-t')\right]dt'$$

and

(11)
$$\langle g^2 \rangle_B \propto \int \left[L_X(t') - L_Y(t')\right]^2 dt' .$$

The sensitivities $S_A$ and $S_B$ to stimuli of type A or B are then defined as

(12)
$$S_A = \sqrt{\langle g^2 \rangle_A} \quad , \quad S_B = \sqrt{\langle g^2 \rangle_B} .$$

To test the effects of adaptation, we measured the filters after the retina was adapted to A – yielding $S_A(A)$ and $S_B(A)$ – and then again after it was adapted to B – yielding $S_A(B)$ and $S_B(B)$. In the course of adaptation to B, the sensitivity $S_A$ changes by a factor $S_A(B)/S_A(A)$, and $S_B$ changes by a factor $S_B(B)/S_B(A)$. The ratio of these two factors is the adaptation index

(2)
$$\alpha = \frac{S_A(B)/S_A(A)}{S_B(B)/S_B(A)}.$$

For the other adaptation experiments, the analysis proceeded in precisely the same fashion. In each case, the probe stimulus spans a broad space that encompasses both stimuli of type A and B. Thus the response filters derived from P allow a measurement of the neuron's sensitivity to stimuli A and B, and ultimately the adaptation index.

## Anti-Hebbian retina model

In the approximation discussed in the text, the retina's instantaneous light response to bipolar cell signals is given by Eqn (3), which becomes in matrix notation

(13) $$\mathbf{y} = (\mathbf{B} + \mathbf{A}) \cdot \mathbf{x} = \mathbf{R} \cdot \mathbf{x}$$

where

(14) $$\mathbf{B} = \left[ b_{ij} \right], \quad \mathbf{A} = \left[ a_{ij} \right], \quad \mathbf{R} = \mathbf{B} + \mathbf{A} = \text{Response matrix}.$$

The bipolar cell synapses $b_{ij}$ are constant, but the amacrine cell synapses $a_{ij}$ evolve as given by Eqn (4), or in matrix notation,

(15) $$\frac{d}{dt}\mathbf{A} = \frac{1}{\tau}\left(-\mathbf{A} - \beta \cdot \left\langle \mathbf{y} \cdot \mathbf{x}^{\mathrm{T}} \right\rangle\right) = \frac{1}{\tau}\left(-\mathbf{A} - \beta \cdot (\mathbf{A} + \mathbf{B}) \left\langle \mathbf{x} \cdot \mathbf{x}^{\mathrm{T}} \right\rangle\right) = -\frac{1}{\tau}\left(\mathbf{A} + \beta(\mathbf{A} + \mathbf{B}) \cdot \mathbf{C}\right),$$

where

(16) $$\mathbf{C} = \left\langle \mathbf{x} \cdot \mathbf{x}^{\mathrm{T}} \right\rangle = \text{stimulus covariance matrix}.$$

Thus the response matrix $\mathbf{R}$ of the network changes according to

(17) $$\frac{d}{dt}\mathbf{R} = -\frac{1}{\tau}\left((\mathbf{R} - \mathbf{B}) + \beta \mathbf{R} \cdot \mathbf{C}\right).$$

After adaptation is complete, $d\mathbf{R}/dt = 0$, and therefore

(18) $$\mathbf{R}(t=\infty) = \mathbf{B}\cdot(\mathbf{1}+\beta\mathbf{C})^{-1}$$

To interpret this response matrix, it is best to work in the eigenbasis of the covariance matrix. $\mathbf{C}$ is symmetric real and therefore has $n$ orthonormal eigenvectors, where $n$ is the number of bipolar cells. Let $\mathbf{u}_j$ denote the $j^{\text{th}}$ eigenvector of $\mathbf{C}$ with eigenvalue $c_j$. In the basis of the $\mathbf{u}_j$, $\mathbf{C}$ is diagonal

(19) $$\mathbf{C} = \begin{pmatrix} c_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & c_n \end{pmatrix}$$

and therefore

(20) $$\mathbf{R}(\infty) = \mathbf{B}\cdot\begin{pmatrix} \frac{1}{1+\beta c_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{1+\beta c_n} \end{pmatrix}$$

So in the final state after adaptation, the system behaves as though a multi-dimensional scaling had been applied to the bipolar cell input: The component of the input vector along the eigenvector $\mathbf{u}_j$ of the covariance matrix gets suppressed by a factor $1/(1+\beta c_j)$. In summary, the system learns to suppress highly correlated components of the stimulus. It does so by subtracting from the ganglion cell input those signals that are effective at predicting it.

The approach to the final adapted state follows a time course with multiple exponentials. Define the deviation from the final state as

(21) $$\mathbf{R}'(t) = \mathbf{R}(t) - \mathbf{R}(\infty)\,.$$

Then Eqn (17) is solved by

$$\mathbf{R}'(t) = \mathbf{R}'(0) \cdot \sum_{j=1}^{n} \exp\left\{ -\frac{1+\beta c_j}{\tau} t \right\} \cdot \mathbf{u}_j \cdot \mathbf{u}_j^{\mathrm{T}} . \tag{22}$$

So the sensitivity of the system along the direction $\mathbf{u}_j$ in stimulus space approaches the final state exponentially with time constant $\tau / (1+\beta c_j)$. Note the approach is faster the higher the stimulus variance $c_j$ along that direction.

In the example of Figs 5d-e, the stimulus drives a 4x4 array of bipolar cells, which are connected to a single ganglion cell as described above. The fixed synapses connect only to the central 2x2 bipolar cells, with equal strength. This default receptive field **R** is apparent when the stimulus is off ($0 < t < 5\tau$). With the appearance of patterned stimulation, the modifiable synapses gradually adjust according to the above rules to suppress the correlated components. For example, stimulation with a flickering grating results in a receptive field with preferred orientation orthogonal to that of the grating ($20\tau < t < 30\tau$). For this illustration, the gain factor $\beta$ was set to 5.

# Supplementary Figure Legends

## Fig S1

Spatial layout of multi-variable flicker stimuli.

## Fig S2

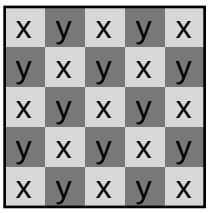
A Linear-Nonlinear cascade to model a neuron's firing rate in response to two stimulus inputs. The inputs $x(t)$ and $y(t)$ are each passed through a linear filter, with impulse response $L_X(\tau)$ and $L_Y(\tau)$ respectively. The results are summed and transformed by an instantaneous nonlinear function $N(g)$ to yield the firing rate $r(t)$.

## Fig S3

Inhibitory synapses are essential for pattern adaptation. Adaptation index measured with horizontal and vertical gratings (as in Fig 2e), in normal Ringer's solution (left) and after addition of 10 $\mu$M Strychnine and 100 $\mu$M Picrotoxin (right) to block inhibitory transmission via glycine and GABA, respectively. In the drug condition, the average adaptation index is not significantly different from 1 (p=0.19).

Fig S1

a


x y x y x
y x y x y
x y x y x
y x y x y
x y x y x

b

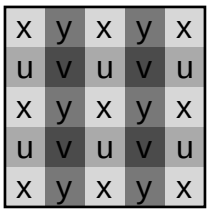

x y x y x
u v u v u
x y x y x
u v u v u
x y x y x

Fig S2

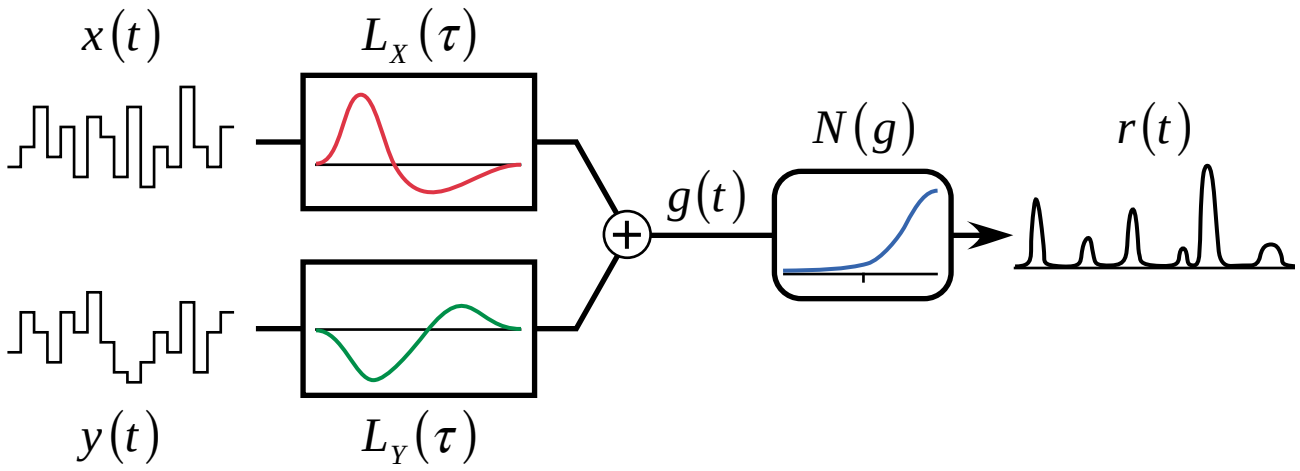
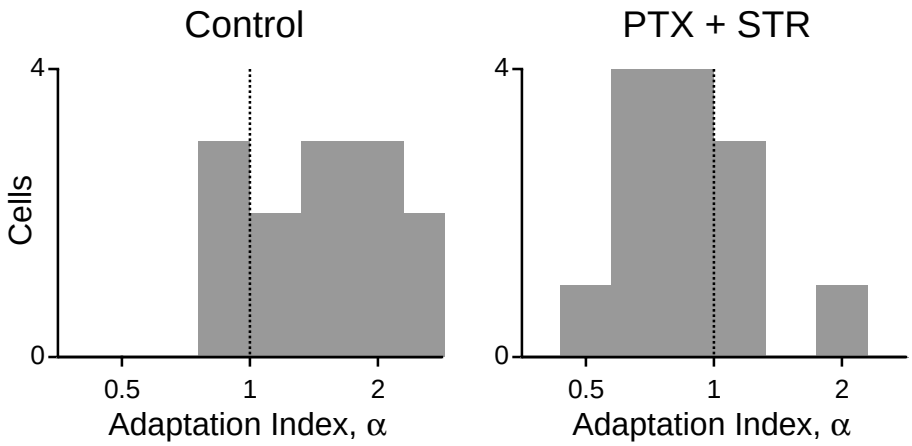


Fig S3


Control
PTX + STR
Cells
4
0
0.5
1
2
Adaptation Index, α
4
0
0.5
1
2
Adaptation Index, α